March 19, 2004

White Paper

On

Evaluation of Sampling Design Options for the National Children's Study

by

Warren Strauss, Jeff Lehman, John Menkedick, Louise Ryan[1], Tim Pivetz, Nancy McMillan, Ben Pierce, and Steve Rust

BATTELLE
505 King Avenue
Columbus, OH 43201-2693

With Critical Review and Input from
Colm O'Muircheartaigh[2] and Alan Zaslavsky[3]

Contract No. 282-98-0019

Dr. Peter Scheidt
Project Officer

Dr. Jim Quackenboss
Work Assignment Officer

National Children's Study Program Office
National Institute for Child Health and Human Development
6100 Executive Blvd - 5C01
Rockville, MD 20892

[1] Louise Ryan is a Professor of Biostatistics at the Harvard University School of Public Health and Dana Farber Cancer Institute.

[2] Colm O'Muircheartaigh is a Professor at the Harris Graduate School of Public Policy Studies at the University of Chicago and Vice President for Statistics and Methodology at NORC.

[3] Alan Zaslavsky is a Professor of Statistics in the Department of Health Care Policy at Harvard Medical School.

DISCLAIMER

This report is a work prepared for the United States Government by Battelle. In no event shall either the United States Government or Battelle have any responsibility or liability for any consequences of any use, misuse, inability to use, or reliance on the information contained herein, nor does either warrant or otherwise represent in any way the accuracy, adequacy, efficacy, or applicability of the contents hereof.

ACKNOWLEDGEMENTS

The authors of this report wish to acknowledge the contributions of various people from different organizations who contributed to the design effort. From the NICHD Program Office, we wish to thank Peter Scheidt and Jim Quackenboss for their technical guidance and assistance throughout the project. We also wish to thank Mark Klebanoff (NIH), Pauline Mendola (US EPA), and Frank Speizer (Harvard) for their input and guidance. From Battelle, we wish to thank Vince Brown, Cody Hamilton, Pamela Hartford, Nicole Iroz-Elardo, Jyothi Nagaraja, Nancy Reiches, Rachel Sell, Lorraine Sinnott, Shannon Stetzer, and Amy Thomas for their technical contributions, and Allison Marshall for her secretarial support. Finally, we wish to acknowledge the input that the NCS Sampling Design Workshop Committee had on this effort, whose members represent the NCS Advisory Committee (Robert Michael, Univ. of Chicago; Barry Ryan, Emory), Study Design Working Group (Jonas Ellenberg, Westat; Frank Speizer, Harvard), Interagency Coordinating Committee (Jim Quackenboss and Sherry Selevan, US EPA), and the NCS Program Office (Peter Scheidt, NICHD; Lester Curtin, NCHS).

Table of Contents

Glossary of Terms

Analytic Study: a study in which action will be taken on a process or cause-and-effect system with the aim of improving future conditions.

Attrition: typically refers to the case where a member of a longitudinal study drops out of the study.

CDC: the U.S. Centers for Disease Control and Prevention.

Centers: a purposively selected medical center capable of performing data collection activities for the NCS – most likely selected through a competitive Federal procurement process.

Certainty Strata: any subset of the study population that can be enumerated which is selected with certainty (weight=1) in a multistage probability-based sampling approach.

Cohort: a group of subjects that are studied over a period of time as part of a scientific investigation.

Confounding: occurs when two factors are associated with each other or "travel together" and the effect of one is confused with or distorted by the effect of the other.

Core Hypotheses: a series of specific research hypotheses deemed by the ICC as sufficient to support the determination of sample size and design for the NCS and essential to assure that specific research questions can be addressed by the study.

Contract research organization: any organization that may be hired through competitive bids (e.g., universities, nonprofit organizations, hospitals, commercial research corporations, etc.) to perform a scope of work for the NCS.

Convenience sampling: a nonprobability sampling approach that selects members based on convenience.

Covariate: a variable that is related to, or has influence on, an outcome of interest.

Cluster Sampling: a method of sampling in which, at some stage, elements (e.g., children) are selected from the population in groups or clusters. In multistage cluster sampling, a sample of elements within a selected cluster may be taken during a subsequent stage of sampling.

Design Effect: a measure of the information loss due to the selected design. Typically defined as the ratio of the parameter estimate variance under a specified sampling design to the parameter estimate variance under a simple random sample.

Design Variables: the set of variables required to implement a probability-based sampling process, including stratification variables and any variables used to calculate probabilities of inclusion.

Effect Modifier: a variable that interacts with a risk factor so that a different association between the risk factor and the outcome of interest is apparent for different values of the effect modifier.

Enumerative Study: a study in which action will be taken on the elements in the frame studied where the term frame is used to refer to an aggregation of identifiable units, any of which may be studied.

External Validity: relationships identified in a study are considered to be externally valid if they are valid for the reference population associated with the study.

EPA: the U.S. Environmental Protection Agency.

Exposure: in this work, exposure is broadly defined as physical, chemical, biological, and/or psychosocial influences that may be related to adverse health outcomes.

GEE: Generalized Estimating Equations – a statistical modeling approach that allows for analysis of correlated data under the conceptual framework of generalized linear models (such as logistic regression models).

Generalize: refers to the ability to draw general conclusions relevant to some population (e.g., apply conclusions to the reference population).

ICC: the interagency coordinating committee – Investigators from each of the four lead agencies (NICHD, CDC, EPA and NIEHS) serve on an Interagency Coordinating Committee (ICC) that is charged with leading the planning and implementation of the NCS.

Inference: a conclusion drawn from evidence.

Internal Validity: relationships are considered to be internally valid if they are statistically significant for the study sample, if the effects of extraneous variables, plausible confounders, and plausible effect modifiers have been properly accounted for, and if hypothesized causal factors precede the effect.

Logistic Regression Model: a statistical analysis method used to model binary or binomial response variables. Parameter estimates from logistic regression models carry log-odds-ratio interpretation.

Model-based Analysis: refers to an inference procedure that implicitly assumes the sampling mechanism does not depend on the survey outcomes.

MSA: Metropolitan Statistical Area.

Multistage Sampling: multistage sampling methods allow selection of groups of elements from the sampling frame at one stage and then subsequent sampling from the selected groups of elements at a subsequent stage.

NCS Cohort: the *study sample* for the National Children's Study.

NCSAC: National Children's Study Advisory Committee (NCSAC), chartered under the Federal Advisory Committee Act, serves as the formal mechanism for providing advice and recommendations to the ICC.

NIEHS: the National Institute for Environmental Health Sciences.

NICHD: the National Institute for Child Health and Human Development.

Non-coverage: refers to the inability to completely identify or enumerate the reference population.

Nonprobability sampling: sampling from the population in some nonrandom manner (i.e., not all members of the population have a known non-zero probability of selection).

Nonresponse: occurs when a member of the population is selected as part of the sample, but, for whatever reason, does not become a participating member of the sample (e.g., a selected person refuses to participate in the study).

NPBS: National Probability-Based Sample.

Odds Ratio: a statistical measure of association. In the context of the design work presented in this report, it is a measure of the relationship between an adverse health effect and a binary measure of exposure. Specifically, it assesses the odds of disease among exposed individuals divided by the odds of disease among unexposed individuals.

Population of interest: could also be called the reference population or the target population (i.e., the population of subjects or units that are the target of the investigation). Typically, inference and/or conclusions are targeted at the population of interest.

Power: probability of correctly concluding that there is an effect when an effect of specified size is present.

Power Studies: studies involving calculation of power under different scenarios.

Probability-Based Random Sampling: a probability-based sampling method for which each element has a probability of being included in the target sample that is strictly greater than zero and strictly less than one, and that uses a random procedure to select elements into the *target sample* according to these probabilities.

Probability-Based Sampling: a method for selecting a *target sample* from a *sampling frame* in which the probability of occurrence for each and every possible *study sample* is a function of a set of *design variables*; an important property of a probability-based sampling process is that the probability of inclusion in the *target sample* is known for each and every element (e.g., child) in the *sampling frame*.

Proportional to Size Sampling: sampling of units with probabilities proportional to the unit size.

PSU: Primary Sampling Unit.

Purposive sampling: nonprobability sampling with some purpose in mind (e.g., purposely sampling a portion of the population that has previously been representative of the population).

Quota Sampling: a method of sampling in which certain characteristics of potential study participants are measured and participants are included in the *study sample* in such a manner as to obtain pre-determined numbers of participants in specified classes defined by values of the measured characteristics.

Recruitment Rate: the ratio of the number of subjects initially enrolled in the NCS cohort divided by the number of subjects for which a recruitment attempt is made.

Reference Population: the population about which valid inferences are desired and to which study inferences will be extrapolated in one form or another.

Representative: used in the context of a representative sample and generally meaning that the sample is "similar to" the population from which it is selected.

Response Rate: the ratio of the number of cohort members providing sufficient data for a particular line of inquiry divided by the number of cohort members for which an attempt is made to collect such data.

Retention Rate: the ratio of the number of actively enrolled cohort members at a given point during the data collection phase of a study divided by the number of cohort members initially enrolled.

Sampling Frame: that portion of the *study population* that has a positive probability of being included in the *target sample*; in practice, the sampling frame is constructed to be as close to the *study population* as possible subject to the requirements that (1) the sampling frame can be fully enumerated and (2) *design variable* values are available for each element of the sampling frame.

Sampling unit: refers to the elements or units that are to be sampled.

Sample Weights: refers to the number of elements/units that are represented by the observation and is typically defined as the inverse of the sampling probability.

Selection bias: a systematic tendency on the part of the sampling procedure to exclude or include one (or more) type(s) of study subjects from the sample.

Simple Random Sampling: simple random sampling methods select the target sample from the sampling frame in a totally random fashion without replacement.

Stratified Sampling: stratified random sampling methods control the subsample sizes for subsets (strata) of the sampling frame defined by one or more design variables.

Study Population: the population of elements that would be included in the *sampling frame* if full enumeration of the sampling frame and values for the design variables were not required.

Study Sample: all elements of the study population that are successfully recruited into the study, are successfully retained as study participants, and produce the required study data.

Target Sample: those elements of the study population for which a recruitment attempt is made; the target sample is the union of the study sample, the set of recruitment failures, the set of retention failures, and the set of retained study participants that fail to produce the required data.

Validation sample: a small sample that is designed to provide information related to the bias or error introduced into the main cohort by nature of the design. The information gathered from the validation sample is designed to allow for appropriate statistical adjustments to the data collected in the larger cohort to address bias and error.

Weighted Analysis: an analysis procedure that appropriately accounts for the sampling weights assigned to each observation.

1 INTRODUCTION

The National Children's Study (NCS) will study the complex relationship between health and the environment for approximately 100,000 U.S. children and their families. Enrollment will begin before birth and follow-up will continue for at least 21 years. Planning for the NCS was initiated by the President's Task Force on Environmental Health Risks and Safety Risks to Children, which was established in 1997. The Task Force was charged with developing strategies to reduce or eliminate adverse effects on children caused by environmental exposures. However, the Task Force soon recognized that such strategies required a much clearer understanding of risk factors, and therefore proposed a longitudinal cohort study of the effects of environmental exposure on the health and development of children (Branum et al., 2002). Title X of the Children's Health Act of 2000 subsequently authorized the National Institute of Child Health and Human Development (NICHD), in collaboration with the Centers for Disease Control and Prevention (CDC), the U.S. Environmental Protection Agency (EPA), and other appropriate Federal agencies, to plan, develop and implement the study.

1.1 NCS SCOPE, OBJECTIVES, GUIDING PRINCIPLES, AND GIVENS

The language in the Children's Health Act of 2000 (Title X, Section 1004) calls for "a national longitudinal study of environmental influences (including physical, chemical, biological, and psychosocial) on children's health and development." The additional direction in the legislation is sparse but critically important. It calls upon the Director of NICHD to "establish a consortium of representatives from appropriate Federal agencies (including the Centers for Disease Control and Prevention, and the Environmental Protection Agency) to (as quoted in subsection (b) of Section 1004):

(1) plan, develop and implement a prospective cohort study, from birth to adulthood, to evaluate the effects of both chronic and intermittent exposures on child health and human development; and

(2) investigate basic mechanisms of developmental disorders and environmental factors, both risk and protective, that influence health and developmental processes.

Finally, the legislation requires that the study shall (as quoted in subsection (c) of Section 1004):

(1) incorporate behavioral, emotional, educational, and contextual consequences to enable a complete assessment of the physical, chemical, biological and psychosocial environmental influences on children's well-being;

(2) gather data on environmental influences and outcomes on diverse populations of children, which may include the consideration of prenatal exposures; and

(3) consider health disparities among children which may include the consideration of prenatal exposures."

The five legislative statements quoted above provide the overall objectives for the NCS.

The legislation and its requirements and their interpretation by the responsible government agencies lead to a set of basic requirements or assumptions for the NCS, which have been referred to as "givens" for NCS sampling designs in the past by government study leaders. These include:

(a) The study will be observational in nature and will address multiple environmental influences.

(b) The study will be national in scope, but not necessarily nationally representative. The sample should be broad-based, inclusive of a wide range of populations and geographic diversity, and as representative as possible given tradeoffs with other features of scientific value to the study objectives. The primary purpose of the study is to investigate exposure-response relationships, not to provide estimates of disease and exposure incidence and prevalence.

(c) The study will include a large sample (approximately 100,000) – to allow for evaluation of rare exposures and outcomes; and of interaction of environmental factors and genetics.

(d) The study will include prenatal recruitment, as early in pregnancy as possible.

(e) The study will include clustering of samples to allow for efficient collection of exposure and outcome measures, and measurement of context (physical and social).

(f) The study will consider stratification to obtain a) an adequate range of exposures (including social), b) socioeconomic, racial/ethnic/geographic diversity, and c) population subgroups of interest.

(g) The study will have locality-based aspects to encourage community engagement.

(h) The study will include infrastructure to support specialized measures (e.g., medical facilities with technologies such as 3D ultrasound).

(i) The study will provide access/collection of appropriate specialized measures or biological samples during pregnancy and birth, for example, placenta or cord blood samples from the delivery room.

(j) The study will provide flexibility to conduct special studies (e.g., special population groups, preconception recruitment, or topics of community interest).

The distinguishing features of the NCS – what makes the study an unusual if not unique research opportunity – are its size (100,000 children), its duration (prenatal, and most likely for a subgroup, peri-conceptional, to adulthood) and its comprehensive charge to assess multiple effects on diverse populations. The legislative requirements that translate to study objectives, and the "givens" stated in terms (a) – (j) above provide the overall boundaries and the guiding principles for the study design. Within these boundaries the overarching goals of the NCS articulated by the Interagency Coordinating Committee (ICC) are to:

- Identify the presence or absence of adverse effects from environmental exposures of concern to development
- Identify possible causal environmental factors for various conditions and developmental and health problems in children and adults
- Provide valuable resources for additional, future studies of health and environment.

1.2 ORGANIZATION AND GOVERNANCE

The Children's Health Act stipulated that the study be carried out with participation of the multiple Federal agencies concerned with children's environmental exposures and possible outcomes. Since fiscal year 2000, a number of interagency agreements have been put into place to carry out methods development studies, provide support services, and establish collaborations among the agencies (NCS Business Plan, 2002). In an effort to solidify this partnership, the four lead institutes and agencies (NICHD, NIEHS, CDC and EPA) signed a Memorandum of Understanding in February 2002. Investigators from each of these four lead entities serve on an Interagency Coordinating Committee (ICC) that is charged with leading the planning and implementation of the NCS, which is coordinated through an NCS program office established at NICHD. By legislative directive, the director of NICHD has overall responsibility and accountability for conduct of the study.

An NCS Advisory Committee (NCSAC), chartered under the Federal Advisory Committee Act, serves as the formal mechanism for providing advice and recommendations to the ICC. The NCSAC is supported by more than 20 Working Groups representing both Federal and private-sector scientists and other specialists focused on providing input on specific scientific questions and issues encountered in study design. In addition, any interested parties receive information on the study and provide input through regularly scheduled Assembly meetings. The overall structure of the NCS leadership is illustrated in Figure 1-1.

Additional information on the history of the evolution of the NCS is available in the following references: Branum et al. (2002), Children's Health Act (2002), NCS Business Plan (2002).

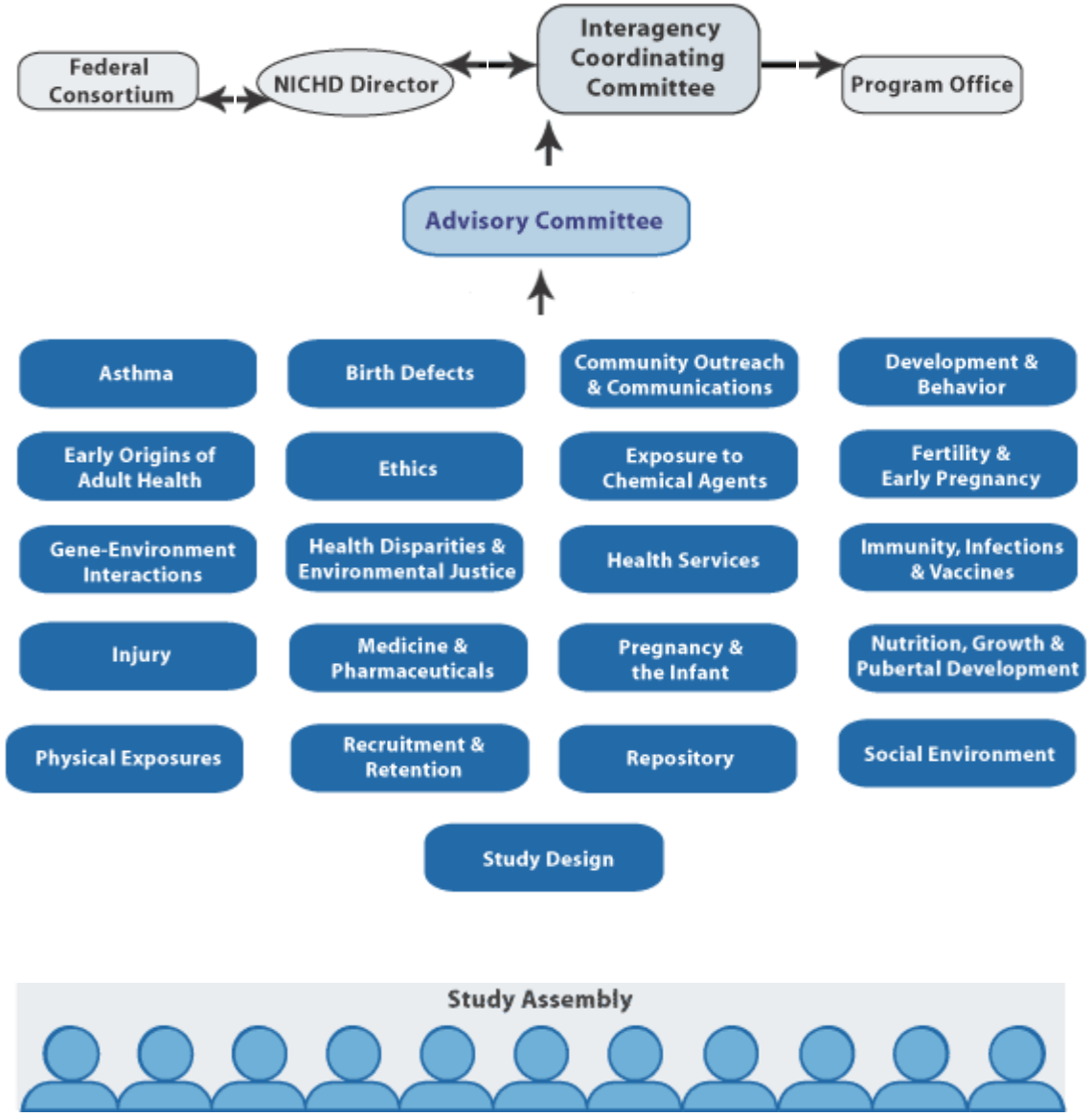

Figure 1-1. NCS Organizational Structure

1.3 SELECTION OF CORE HYPOTHESES

The italicized text that follows on selection of core hypotheses was taken from a November 25, 2003, document prepared by the ICC hypotheses subcommittee, and presented to the NCS Advisory Committee on December 15, 2003.

> *Hundreds of scientists and representatives from community groups and professional organizations have contributed to the identification of key children's environmental health questions. No single research question is of sufficient breadth or import to fulfill the entire mission of the NCS. The Study Design Working Group of the NCS Advisory*

Committee (NCSAC) proposed the development of core hypotheses encompassing exposures and child health outcomes of great public health significance requiring long-term follow-up and which cannot be reasonably studied with fewer children or a different study design. The set of research questions forming the foundation of the NCS must together provide: a rationale for a long-term, prospective study of approximately 100,000 children; the scientific framework to define the NCS, including sample design, data collection, etc.; as well as a "public identity" for the NCS.

The Interagency Coordinating Committee (ICC) has used the findings from 20 NCS working groups reported via the NCSAC, independent reviews of the children's environmental health literature, and comments from a broad-based Study Assembly to develop an initial set of these foundational, core hypotheses. These hypotheses are sufficient to support the determination of sample size and design for the NCS and are essential to assure that specific research questions can be addressed by the study.

However, a manageable set of core hypotheses cannot alone convey the true breadth of the NCS, nor do they, alone, assure the collection of data necessary to address the full range of topics to be covered by the NCS. The priority outcomes and exposures outlined below go further to convey the full scope of the NCS. Additional work is necessary to complete a study protocol that balances participant and family burden with data collection activities needed to address these important areas of children's environmental health.

> *<u>Priority Outcomes</u>. Based on the above criteria, the following child health areas have been identified as priorities for the NCS.*
>
> *Pregnancy outcomes: Many pregnancy outcomes, including preterm delivery and birth defects, are plausibly related to environmental conditions and are understudied. These early life events can have profound impact on child health and development throughout life. These outcomes also provide a first set of results from the NCS right from the start.*
>
> *Neurodevelopment and behavior: Assessment of child development and behavior is key to the mandate of the NCS. The NCS can address multiple environmental factors that are potentially associated with severe health concerns such as autism and schizophrenia, as well as more commonly occurring childhood disorders such as depression and learning disabilities. The NCS can also provide substantial data on variations in the course of normal child development and may provide insights into environmental factors related to aspects of development such as aggression, adjustment, achievement and resilience.*
>
> *Injury: A focus area of the President's Task Force, injury is a major cause of childhood morbidity and mortality. The NCS expects to measure childhood injuries, particularly those that require hospitalization or other medical attention, and to evaluate a variety of environmental factors including aspects of the social and physical environment that may be associated with injury.*

Asthma: While there is a substantial body of research into environmental factors that can trigger asthma attacks or exacerbate existing asthma, there is a need to understand more about contributions the environment and gene-environment interactions have on the development of asthma. Because asthma is relatively common among U.S. children, the NCS will have enough statistical power to be able to examine various constellations of environmental and genetic factors that may be related to asthma incidence and exacerbation.

Obesity and physical development: The NCS will likely have sufficient statistical power to examine disorders of physical development related to diabetes, obesity and altered puberty. The longitudinal nature of the data and the ability to examine the interaction of multiple environmental factors with an individual's genetic composition is expected to provide insight not only into growth-related disorders, but also to provide a strong study of variations in growth, physical and reproductive development that may be affected by the environment.

Priority Exposures. The priority exposures listed below are outlined by influence (either beneficial or deleterious) on child health and development:

Physical environment: The NCS will measure aspects of the physical environment, including housing quality and neighborhood and community conditions that may relate to child health and development. In addition, the influence of physical factors such as radiation (electromagnetic, ultrasound, microwave, x-irradiation), light, and noise may be studied.

Chemical exposures: Exposure to chemical environmental contaminants generally occurs through human contact with air, water, soil, dust, food or industrial products. Pollutant exposures currently of concern in the NCS include metals, PCBs and dioxins, phthalates, organic and inorganic pesticides and herbicides. Exposure to many of these compounds, and their mixtures, at low background levels is ubiquitous. The NCS can investigate the potential health effects associated with these complex low-level exposures. Additionally, the NCS may select specific populations with unique exposure scenarios for special sub-studies of related health effects.

Biologic environment: The biologic environment includes exogenous factors (e.g., infectious agents, endotoxin, diet) and individual response to those factors (e.g., inflammatory response, glucose metabolism). In utero and early life exposures have potential implications for a wide range of health conditions including birth outcome, developmental outcomes, asthma, obesity, and cardiovascular disease. The NCS will allow for elucidation of those associations as well as physiologic mechanisms underlying those relationships, including the influence of genetic composition on those interactions.

Genetics: The NCS offers a unique opportunity to investigate the genetic component of many health outcomes. Although it is recognized that genetic factors play a role in many conditions, the mechanism behind the genetic contribution to specific diseases, such as autism, remains unknown. In addition, the quantitative contribution of genetics to more general conditions, such as obesity, is also unknown. A complete understanding of the effects of the environmental factors listed above requires elucidation of the interactions between these factors and genes, including the roles played by various polymorphisms in environmentally responsive genes and the effects of exposures on gene expression. The large sample size will allow for examination of the interaction between genetic make-up and chemical, biologic, and social exposures on many outcomes. The longitudinal and prospective nature of the NCS offers the possibility of examining the potential development of somatic mutations in relation to specific exposures. The current state of the science likely does not allow for the genetic profiling of study participants but will, at least initially, require a focus on suspect candidate genes. This will change as the study matures.

Psychosocial milieu: The NCS expects to assess many potential aspects of the psychosocial environment including: families and households; socioeconomic status; social networks and social support; neighborhoods and communities; formal institutions; and public policy. These factors have the potential to influence a child's health either directly or indirectly, by affecting exposure to the chemical or physical environment. The NCS will be able to examine those associations as well as shed light on the physiologic mechanisms underlying, for example, potential relationships between psychosocial stress and asthma or preterm birth. In addition to the putative influence on the health of an individual, social environmental factors may be an important area of consideration for investigation of health disparities.

Integrating Priority Outcomes And Exposures. Based on input from hundreds of experts, the ICC has proposed a set of core hypotheses to define a framework for study design. Though the current list of core hypotheses [see Table 6.1 in Chapter 6] *is still under debate, it is largely accepted as being adequate to move forward with development of a sampling design. It is expected that, over the long course of the study, new questions will emerge and be added to the study and some of the core hypotheses here may become outdated (ICC Hypotheses 2004).*

1.4 GUIDANCE FOR THE STUDY DESIGN

The ultimate purpose of the study design is to define all study specifications – cohort selection, measurement specification, and implementation details – in a manner that will best meet the overall objectives, requirements, and goals of the study described in Section 1.2 above. In addition, more specific guidance has been articulated (by government study leaders, the Study Design Working Group, and the NCSAC) that calls for the study design to:

- Emphasize hypotheses and science needs that require the unique longitudinal nature or sample size of the NCS;
- Go beyond characterization of associations, to provide understanding of the causal relationship between exposure and disease;
- Cover a sufficient range of exposures and outcomes to understand significant interactions;
- Be as representative as possible of the U.S. population, with relationships between exposure and disease able to be generalized to a broader population;
- Provide sufficient power to detect target associations of interest for selected core hypotheses;
- Provide a resource to test hypotheses to be identified in the future;
- Allow assessment, as possible, of populations at higher risk of exposures or outcomes;
- Be transparent, with assumptions, tradeoffs, and decisions well-documented; and
- Address ethical considerations, including cohort burden.

The difficulty (or challenge) in meeting the goals for the study design lies primarily in two areas – first, the fact that design choices must be made in the face of scientific and implementation uncertainties, and second, the fact that even with a good understanding of what might be expected there are tradeoffs between conflicting objectives.

The most notable example of the difficulty in meeting multiple goals for the study design is the ongoing difference in opinions over the feasibility and desirability of certain aspects of probability sampling. On the one side, many epidemiologists believe that a strict probability approach will result in fewer measurements, more attrition, and negative impact on the ability to measure exposures and outcomes sufficiently well to understand the etiology of disease. On the other hand, sampling statisticians and social scientists are concerned that the lack of probability sampling may introduce unknown biases into study results and leave the study with results that cannot be generalized to a broader population. This controversy arises first because there is uncertainty over the degree to which a probability sample will result in more attrition and less measurement in comparison to a convenience sample; and second because there are tradeoffs involved between maximizing a probability component to the sample and many other desirable features such as efficiency and accessibility of measurements, local community involvement, and support for major research institutions. The white paper on the Advantages and Limitations of Probability-Based Sampling for the National Children's Study included in Appendix A and the two white papers on Criteria and Design Options included in Appendices B1 and B2 provide more discussion on the specific tradeoffs and uncertainties associated with study design choices in the NCS.

It is for this reason that study leaders have convened the sampling workshop to discuss tradeoffs, and identify a study design approach that maximizes advantages and minimizes disadvantages, in light of uncertainties and conflicting objectives.

1.5 HISTORY OF THE DESIGN EFFORT

Work on issues associated with the study design for the NCS began with the creation of the ICC. Members of the ICC, NICHD program staff, and members of the NCSAC have engaged in a rich discussion of options and possibilities. In addition to this ongoing dialogue, there were three directed efforts at preparing for a study design that deserve particular mention.

The first (and ongoing) effort is the contribution of the Study Design Working Group. Beginning in 2001, the Study Design Work Group has met, discussed study design needs, provided findings through the Advisory Committee to the ICC and Program Office, and requested pilot studies necessary to help inform design decisions. The Working Group originally focused on helping identify candidates for core hypotheses for the study and the criteria that might be used to judge candidate hypotheses. Later the Working Group focused on review of sampling designs proposed in the Westat report discussed below. This included comments and findings provided through the NCSAC to the ICC. The Sampling Workshop Planning Committee that planned the March 2004 NCS Sampling Workshop includes two members from the Study Design Working Group who continue to provide input from this working group.

The second effort is a report prepared by Westat, under contract to the National Center for Health Statistics, with guidance from members of NICHD and NCHS. The purpose of the Westat report was to develop and evaluate a number of candidate sample frames and sample designs for NCS enrollment. The report discusses three sampling models for initial consideration: a Household Model (door-to-door screening for fecund women), an Office Model (recruitment of pregnant women during ordinary prenatal care visits), and a Center Model (recruitment of pregnant women through a small number of formal centers that would be responsible for executing all aspects of the study protocol for their own recruits throughout the life of the project). Two variants of the Household Model with different degrees of clustering were examined, resulting in evaluation of four candidate designs. The report discussed the type and degree of clustering in the four evaluated designs, initial sample size determination, detailed costs for the sample recruitment, some aspects of the relative difficulty of various measurements of exposure and outcomes under the alternative designs, and statistical power for various tests (Westat, 2002). The Westat report significantly advanced the study design effort by providing detailed candidate options for consideration.

The third effort is a report prepared by Battelle, under contract to the U.S. Environmental Protection Agency, that examined optimal design considerations for measuring environmental exposures in the NCS, including methods for improving estimates of exposure through the use of detailed sub-studies that collect more precise exposure information on a small validation subsample of participants and use latent variable models to assess the relationship between health outcome and environmental exposure in the presence of measurement error. The methodology presented in the Battelle report is relevant to the overall sampling design in that it provides a tool that can reduce burden across the cohort and therefore potentially impact the feasibility of different sampling designs. A summary of the Battelle report to EPA is provided in Appendix C.

In September 2003, the NCS Program Office contracted with Battelle to prepare a white paper (Appendix B2) outlining a range of design options for selecting the longitudinal cohort into the study, building off the sampling design work described above. The Battelle paper first discussed options for three primary design elements which were seen as fundamental aspects of any proposed design. The design elements were: choice of the sampling frame for the population, method of selecting participants for the cohort, and organizational structure of the study.

For the first design element, the options paper presents three primary candidates for a sampling frame, largely synonymous with those presented in the Westat report. The first candidate was a household sampling frame that consists of a set of identifiable households in the U.S., and operationally would involve screening a sample of households to identify pregnant women, women of childbearing age, and/or couples attempting pregnancy. The second was a physician's office sampling frame which would allow for the selection of a sample of physicians and/or medical offices during a first stage of sampling, and the recruitment of a sample of pregnant women and/or women of childbearing age seen in their practices during a second stage of sampling. The third candidate was a community or university medical center sampling frame that involves selecting a sample of large health centers during the first stage of sampling that have previously demonstrated their ability and interest in conducting the NCS data collection protocol (e.g., through a competitive proposal process). These centers would recruit pregnant women and/or women of childbearing age either in proximity to or currently being served by their center or associated physician's offices.

The second design element discussed in the Battelle options paper addressed the methods for sampling the cohort of subjects from the sampling frame (i.e., selecting the subjects that will participate in the NCS). The range of options for selecting the cohort began with a set of fundamentally simple sampling design options that result from a choice of whether (1) the Primary Sampling Units (PSUs) are selected via probability-based sampling, quota methods, or some type of other non-probability method, and (2) participants within the PSUs are chosen via probability-based sampling, quota methods (to ensure some diversity and/or some similarilty with the larger population), or some other type of non-probability-based method. In addition to these fundamentally simple sampling designs, a class of hybrid design options was described. These hybrid design options combine probability-based sampling and non-probability-based sampling by selecting a portion of the sampling units on a probability basis and selecting all other sampling units on a quota or other non-probability basis, both for PSU selection, as well as for selection of participants within a PSU. The methods for specifying hybrid options for cohort selection introduced in the initial options paper (Appendix B2) represents a starting point, with subsequent development of a framework for the family of designs presented for consideration in Chapter 3 of this report.

Finally, the last design element covered in the options paper was the choice of an organizational structure for conducting the NCS and implementing the data collection protocols. The options for the organizational structure discussed included primarily University medical centers or large hospitals, contract research data organizations, health care providers, or some combination of the three.

After discussion of the design elements, the Battelle options paper discussed six general design categories or classes for recruiting and retaining the NCS cohort and the advantages and disadvantages of each. These included:

1. Complete probability-based design (all units at all levels are selected on a probability basis).
2. Convenience or quota sampling of PSUs and within PSU probability-based sampling.
3. Complete convenience or quota sampling.
4. A combination of convenience and probability-based sampling of PSUs, and complete probability-based sampling within PSUs.
5. A combination of convenience and probability-based sampling of PSUs and within PSUs.
6. A multiple cohort design with convenience selection of one (or more) cohort(s) and probability-based sampling of another (or other) cohort(s). The multiple cohorts could undergo varying levels of data collection (e.g., less burdensome environmental, behavioral, and health outcomes sampling for the probability sampled subjects), and could be followed for varying periods of time.

This Battelle options paper, along with a companion paper on criteria for evaluating the design options (see Appendix B1), served as the basis for discussions between NICHD Program Office representatives, Battelle staff members, and two consultants, Dr. Alan Zaslavsky of Harvard Medical School, and Dr. Colm O'Muircheartaigh of the University of Chicago's Harris School of Public Policy Studies. The complete summary of these meetings is included in Appendix B3. As discussed in the following Section, the final outcome was consensus agreement to explore a family of designs rather than pursue purely probabilistic or non-probabilistic designs.

1.6 THE FAMILY OF DESIGNS

At the Battelle meeting, all participants acknowledged that both the probability-based selection approach and the non-probability-based selection approach offer advantages and disadvantages, and both approaches have certain limitations in light of the objectives and constraints of the NCS. As discussions progressed, the meeting participants began to share the opinion that both of these sample selection methods offer important components to the NCS and may be able to be accommodated in the design. The group recognized that different categories of study users had legitimate scientific objectives that would favor probability sampling in some instances and restrictions on probability sampling to achieve other scientific objectives in other instances. For example, probability-based sampling offers the ability to generalize the results of the study with minimal assumptions; however other types of sampling approaches might offer more flexibility in obtaining previously collected medical history information from a more narrowly defined subset of potential respondents. Therefore, the group recognized a continuum of sampling methods in which a complete non-probability sample is at one extreme of the continuum and a complete probability-based sample is at the other extreme. Somewhere in the middle of these two extremes (i.e., a design that selects some portion based on probability and

some portion non-probabilistically) may lie an optimal design that can satisfy *most* (ideally all) of the objectives of the NCS. In keeping with the Battelle sampling design options paper, this might be called a "hybrid" design, but it may be better referred to as a *family of designs*. In other words, the NCS may not be composed of a single design, but rather a variety, or family, of designs that can be combined to address the multiple objectives of the NCS.

With respect to the concept of a family of designs, the meeting participants agreed that this type of design would be used to tackle multiple hypotheses and objectives. Thus, different parts of the design would be best suited to service different hypotheses and research demands. Some parts of the design would be essential for measures where data could be collected only in or by major medical centers; other parts of the design would protect against unforeseen circumstances and biases, protect against under-coverage of particular parts of the population that might undermine the validity of an inference, and allow statistical inferences to be extended to the whole population of the U.S. In terms of the application of evaluation criteria to the family of designs, the participants thought that it would be useful to check designs explicitly against criteria such as those proposed in the white paper included in Appendix B. By thinking of a family of designs, however, it is quite possible that a particular member of the family may fail a critical criterion, but may contribute enough on other criteria to make its inclusion not only worthwhile, but essential. Considering the array of designs and the array of criteria jointly, as well as the features and needs of "family members," is what will make the overall design a success.

The discussions also identified other proposed rationales for using a family of designs for the NCS. These rationales are generally related to the size of the study and the ability to propose a design that will meet the objectives of a variety of researchers (medical researchers, epidemiologists, social scientists, health researchers, clinicians, etc.), for whom the values of probability sampling, intensity of data collection, and exposure measures, etc., are of differing relative importance. First, since the sample size for the NCS is so large (100,000), the possibility of splitting the cohort into a portion selected non-probabilistically and a portion selected randomly could result in large sample sizes for both groups of individuals (whereas, in most studies that involve a small cohort of individuals, splitting of the cohort would not produce reasonable sample sizes). Second, since there are a variety of opinions as to the appropriateness and limitations of probability and non-probability-based selection for the NCS, incorporation of both types of sampling through a family of designs may provide a sampling design that can meet the objectives of a variety of NCS stakeholders. Finally, a family of designs might provide adequate coverage of populations that might not be served or included by more limited sampling frames.

1.7 ROADMAP TO THE REST OF THE REPORT

The purpose of the remainder of this report is to provide technical details on the options for a family of designs in sufficient detail to allow the NCS Sampling Workshop participants to make recommendations on the NCS Study Design. The hope is that the design framework presented in this report is sufficiently clear and reasonable to allow recommendations to be made for a study design.

Chapters 2 – 4 outline the options related to sampling frame, selection of the cohort, and organizational structure for implementation. Chapter 2 discusses target populations and the candidate sampling frames that might be chosen to represent those populations. Chapter 3 focuses on candidate methods for selecting participants, further introducing the family of designs concept and terminology. Chapter 4 introduces the candidate organizational structures for implementation.

Chapters 5 – 9 then discuss technical details critical to the evaluation of the options presented in Chapters 2 – 4. Chapter 5 discusses technical details on implementation of various sampling strategies incorporated in the family of designs, and the impact of the choice of a mixture of sampling strategies on the precision of estimates for the relationship between health effects and exposure (the design effects). Chapter 6 reviews core hypotheses and the measures that are critical to testing these hypotheses, providing a basis for the specific hypotheses chosen to be investigated in the power studies. Chapter 7 provides an overview of the assumptions used concerning recruitment and retention. Chapter 8 presents a model for estimating costs associated with the study as well as initial cost estimates and the assumptions on which they are based. Chapter 9 presents the results of analyses to characterize the power of different design options to detect significant associations for selected core hypotheses.

Finally, Chapter 10 summarizes the results, discusses caveats and limitations, and makes recommendations related to future work, including pilot studies that would help inform design decisions.